自然言語処理 111－9
（1996. 1. 19）

動詞多義性解消における格要素の貢献度について

藤井敦, 乾健太郎, 徳永健伸, 田中穂積
東京工業大学大学院情報理工学研究科
{fujii,inui,take,tanaka}@cs.titech.ac.jp

事例ベースの動詞多義性解消について新しい手法を提案する。本手法では、動詞に係る格要素がその動詞の語義弁別に貢献する度合を定量化し、貢献度の高い格要素における入力・事例間の類似度を優先的に評価することによって入力文に最も近い事例を選択する。10種類の動詞について1種類あたり平均約100文の事例を用いて実験した結果、従来の手法の誤り率を約30%低減することが確認された。

To What Extent Does Case Contribute to Verb Sense Disambiguation?

FUJII Atsushi, INUI Kentaro,
TOKUNAGA Takenobu, TANAKA Hozumi
Department of Computer Science
Tokyo Institute of Technology
{fujii,inui,take,tanaka}@cs.titech.ac.jp

This paper proposes a new example-based method for verb sense disambiguation, in which we consider the degree of contribution of cases to disambiguation, and reportes its performance through our experiment on ten Japanese verbs. Accoding to the experiment, our method reduces about 30% of error rate from the previous method.

1 Introduction

Word sense disambiguation is an essential task in Natural language processing. Most methods proposed until now can be roughly classified into two classes: rule-based approaches and corpus-based approaches. Rule-based approaches try to generalize the contexts where a given word can appear in a certain sense, and obtain rules for word sense disambiguation, e.g. subcategrization rules. As many researchers have pointed out (for example, [6]), this approach requires a tremendous amount of overhead to derive the rules, and moreover it severely suffers from exceptions. In contrast, corpus-based approaches, which have recently received the most attention, take into account similarity between the context of an input word and the context in which the word appears in different senses in a corpus. Word sense disambiguation is executed based on an intuitively feasible assumption that the higher the degree of this similarity is, the more plausible it becomes that the word is used in the same sense. Unlike the rule-based approach, this approach frees us from the task of generalizing observed phenomena to produce rules. Corpus-based methods are further classified into two approaches: example-based approaches [5, 11] and statistic-based approaches [1, 2, 8, 9, 12, 13]. We follow the example-based approach in explaining its effectivity for verb sense disamibiguation in Japanese.

A representative example-based method for verb sense disambiguation was proposed by Kurohashi and Nagao (Kurohashi's method) [5]. Their method uses an example database containing examples of collocations as in figure 1. Figure 1 shows a fragment of the entry associated with the Japanese verb *toru*. As with most words, the verb *toru* has multiple senses, a sample of which are "to take/steal," "to attain," "to subscribe" and "to reserve." The database gives one or more case frame(s) associated with the verbs for each of their senses. In Japanese, a complement of a verb, which is a constituent of the case frame of the verb, consists of a noun phrase (case filler) followed by a case marker such as *ga* (nominative) or *o* (accusative). The database has a set of case filler examples for each case. As shown in figure 1, examples of a complement can be considered as an extensional description of the selectional restriction on it.

The task considered in this paper is to interpret verbs in input sentences. In this paper, "to interpret a verb" means the task of choosing one sense from a set of candidate senses of the verb. Given an input sentence, Kurohashi's method interprets the verb in the input by computing semantic similarity between the input and examples. For this computation, Kurohashi's method experimentaly uses the Japanese word thesaurus *Bunruigoihyo* [7]. Figure 2 illustrates a fragment of *Bunruigoihyo* including nouns associated with the nominative and the accusative in figure 1 respectively. Let us take the example sentence (1).

(1)	*hisho ga*	*shindaisha o*	*toru.*
	(secretary-NOM)	(sleeping car-ACC)	(?)

In this example, it may be judged according to figure 2 that *hisho* ("secretary") and *shindaisha* ("sleeping car") in (1) are semantically similar to *joshu* ("assistant") and *hikôki* ("airplane"), respectively, which are examples that collocate with *toru* ("to reserve"). As such, the sense of *toru* in (1) can be interpreted as "to reserve."

According to an experiment on ten verbs, which will be shown later in section 4, Kurohashi's method improves the accuracy rate of verb sense disambiguation from the lower bound of 42.3% gained by using only relative frequency information, to 76.2%. We investigated the errors made by Kurohashi's method and found a few types of frequent error cases.

One typical case is exemplified by the input sentence (2).

(2)	*shachô ga*	*shûkanshi o*	*toru.*
	(president-NOM)	(magazine-ACC)	(?)

The nominative, *shachô* ("company president"), in (2) is found in the "to attain" case frame of *toru* and there is no other co-occurrence in any other sense of *toru*; therefore, the nominative supports an interpretation of "to attain." On the other hand, the accusative, *shûkanhi* ("magazine"), is most similar to the examples contained in the accusative of the "to subscribe" and therefore the accusative supports the interpretation "to subscribe." Although the most plausible interpretation here is actually the latter, Kurohashi's method would choose the former since (a) the degree in which the nominative supports "to attain" happens to be stronger than the degree in which the accusative supports "to subscribe," and (b) their method always relies equally on the similarity in the nominative and the accusative. In fact, in their method, the plausibility of an interpretation of an input verb is computed simply by averaging the degree of similarity between the input complement and the example complements for each case. However, in the case of *toru*, since the semantic range of nouns collocating with the verb in the nominative do not seem to have a strong delinearization in a semantic sense, it would be difficult, or even risky, to properly interpret the verb sense based on the similarity in the nominative. In contrast, since the ranges are diverse in the accusative, it would be feasible to rely more strongly on the similarity in the accusative. This argument suggest the computation of the plausibility of verb sense by the weighted average of the degree of simliarity for each case.

toru:				
suri (pickpocket) *kanojo* (she) *ani* (brother)	*ga*	*kane* (money) *saifu* (wallet) *otoko* (man) *uma* (horse) *aidea* (idea)	*o*	*toru* (to take/steal)
kare (he) *kanojo* (she) *shachô* (company president) *gakusei* (student)	*ga*	*menkyoshô* (license) *shikaku* (qualification) *biza* (visa)	*o*	*toru* (to attain)
kare (he) *chichi* (father) *kyaku* (client)	*ga*	*shinbun* (newspaper) *zasshi* (journal)	*o*	*toru* (to subscribe)
kare (he) *dantai* (group) *ryokôkyaku* (passenger) *joshu* (assistant)	*ga*	*kippu* (ticket) *heya* (room) *hikôki* (airplane)	*o*	*toru* (to reserve)
⋮		⋮		⋮

Figure 1: Showing a fragment of an example database, and the entry associated with Japanese verb *toru*

One may argue that, given a sufficiently large number of examples, say tens of thousands for each verb [13], this effect would automatically emerge. That is, in the case of *toru*, the similarity in the nominative between frames would become less pronounced, and thus would diminish its potential for disambiguation of sense. On the other hand, the simlarity in the accusative would become more salient. However, this would not be the case were the number of examples not large enough. Our claim in this paper is that we need to consider the problem of data sparseness and can partially overcome this problem by introducing the notion of the degree to which each case contributes to verb sense disambiguation. In the following sections, we will propose a method to achieve this aim and will discuss why considering this notion improves Kurohashi's method given sparse example data, showing our results in an experiment.

Another typical problem with Kurohashi's method is that with sparse data, it freqently gives the same highest plausibility score to more than one interpretation of an input verb, which means that the verb sense ambiguity still remains. Let us consider another example (3).

(3)	*onîsan ga*	*omocha o*	*toru.*
	(brother-NOM)	(toy-ACC)	(?)

In (3) the most plausible interpretation of *toru* is "to steal." The nominative does not give much information for interpreting the verb for the same reason as example (2). In the accusative, the databse in figure 1 has two example case fillers that are equally similar to *omocha* ("toy"): *saifu* ("wallet") and *hikôki* ("airplane"). These examples support two different interpretations: "to steal" and "to reserve," respectively. Here, one may notice that the accusative examples in the case frame of *toru* ("to reserve") are less diverse in meaning than the other case frames. If so, it could be said that the selectional restriction on the accusative of *toru* ("to reserve") is relatively strong, and thus that it is relatively implausible for *omocha* ("toy") to satisfy it. If such reasoning is correct, given that the examples in the accusative of *toru* ("to steal") are most widely distributed, the input verb can be interpreted as "to steal." The consideration above motivated us to introduce the notion of relative strength of selectional restriction into our example-based verb sense disambiguation method. In this paper, we will show how the relative strength of selectional restriction can be computed and report the degree to which such considerations contribute to verb sense disambiguation with the results of our experiment.

Section 2 briefly describes our method. Sections 3 elaborates a way to compute the degree of contribution of case to verb sense disambiguation, and the relative strength of selectional restriction. Section 4 reports the results of our experiment. Before concluding in 6, discussion and related work are added in section 5.

2 Overview

We assume that inputs are simple sentences, each one of which consists of a sequence of cases followed by their governing verb. The task is to identify the sense of each input verb. The set of verb senses we use are those defined in the existing dictionary "IPAL" [4]. IPAL also contains example case fillers

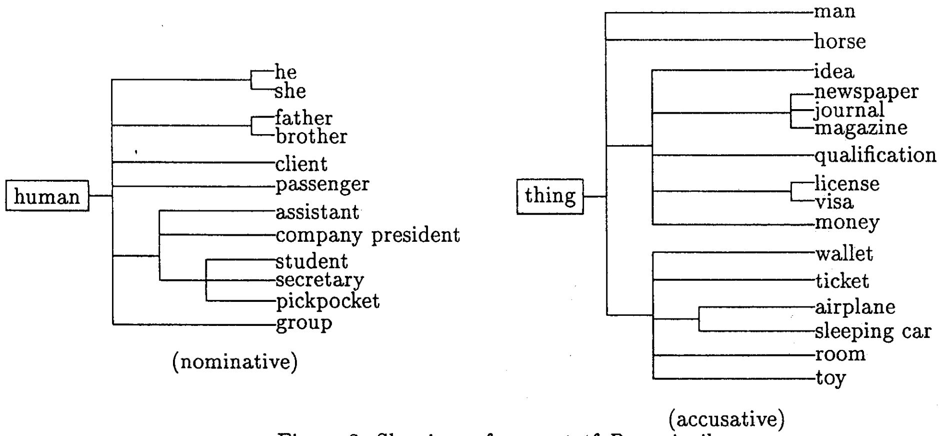

Figure 2: Showing a fragment of *Bunruigoihyo*

Table 1: The relation between the length of path between two nouns (X and Y) in *Bunruigoihyo* and the similarity between them ($sim(X,Y)$)

length of path between X and Y	0	2	4	6	8	10	12
$sim(X,Y)$	11	10	9	8	7	5	0

as shown in figure 1. As well as Kurohashi's method the similarity between two case fillers, or more precisely the semantic-head nouns of them, is computed by using *Bunruigoihyo* [7]. As with most thesauruses, the length of the path between two words in *Bunruigoihyo* is expected to reflect the similarity between them. Following Kurohashi's method, we define $sim(X,Y)$, which stands for the similarity between words X and Y, as in table 1. It should be noted here that both methods are theoretically independent of what resources are used.

To illustrate the overall algorithm, we replace the illustrative cases mentioned in section 1 with a slightly more general case as in figure 3. The input is $\{n_{c_1}\text{-}m_{c_1}, n_{c_2}\text{-}m_{c_2}, n_{c_3}\text{-}m_{c_3}, v\}$, where n_{c_i} denotes the case filler in the case c_i and m_{c_i} denotes the case maker of c_i. The candidates of interpretation for v are derived from the database as s_1, s_2 and s_3. The database also gives a set $\mathcal{E}_{s_i,c_j}$ of case filler examples for each case c_j of each sense s_i. "—" denotes that the corresponding case is not allowed.

In the course of the verb sense disambiguation process, the system first discards the candidates whose case frame constraint is grammatically violated by the input. This is done in both Kurohashi's method and ours. In the case of figure 3, s_3 is discarded because the case frame of v (s_3) does not subcategrize the case c_1[1]. In contrast, s_2 will not be rejected at this step. This is based on the fact that in Japanese, cases can be easily omitted if they are inferable from the given context.

After checking grammatical case frame constraints, the system computes the plausibility of the remaining candidates of interpretation and chooses the most plausible interpretation as its output.

In Kurohashi's method, the plausibility of an interpretation is computed by averaging the degree of similarity between the input complement and the example complements[2] for each case as in equation (1), where $P(s)$ is the plausibility of interpreting the input verb as sense s, and $SIM(n_c, \mathcal{E}_{s,c})$ is the degree of the similarity between the input complement n_c and example complements $\mathcal{E}_{s,c}$.

$$P(s) = w_s \sum_c SIM(n_c, \mathcal{E}_{s,c}) \qquad (1)$$

$SIM(n_c, \mathcal{E}_{s,c})$ is the maximum degree of similarity between n_c and each of $\mathcal{E}_{s,c}$ as in equation (2).

$$SIM(n_c, \mathcal{E}_{s,c}) = \max_{e \in \mathcal{E}_{s,c}} sim(n_c, e) \qquad (2)$$

w_s is the weight on an interpretation s such that the more obligatory cases of those imposed by s are found in the input, the greater the value of the weight is. For more detail, see Kurohashi's paper [5].

In our method, on the other hand, for the reason indicated in section 1, we introduce two new factors:

[1] Since IPAL does not necessarily enumerate all the possible optional cases, the absence of case c_1 from v (s_3) in the figure may denote that c_1 is optional. If so, the interpretation s_3 should not be discarded in this stage. To avoid this problem, we use the same technique as used in Kurohashi's method. That is, we defined several particular cases beforehand, such as the nominative, the accusative and the dative, to be obligatory, and impose the grammatical case frame constraint as above only in those obligatory cases. Optionality of case needs to be further explored.

[2] $\mathcal{E}_{s_2,c_4}$ is not taken into consideration in the computation since c_4 does not appear in the input.

input	n_{c_1}-m_{c_1}	n_{c_2}-m_{c_2}	n_{c_3}-m_{c_3}		v (?)
database	$\mathcal{E}_{s_1,c_1}$	$\mathcal{E}_{s_1,c_2}$	$\mathcal{E}_{s_1,c_3}$	—	v (s_1)
	$\mathcal{E}_{s_2,c_1}$	$\mathcal{E}_{s_2,c_2}$	$\mathcal{E}_{s_2,c_3}$	$\mathcal{E}_{s_2,c_4}$	v (s_2)
	—	$\mathcal{E}_{s_3,c_2}$	$\mathcal{E}_{s_3,c_3}$	—	v (s_3)

Figure 3: An input and the database

- contribution of case to verb sense disambiguation (CCD),
- relative strength of selectional restriction (RSSR).

First, considering CCD, we compute the plausibility of an interpretation by the *weighted* average of the degree of similarity for each case as in equation (3), instead of equation (1).

$$P(s) = \frac{w_s \sum_c SIM(n_c, \mathcal{E}_{s,c}) \cdot CCD(c)}{\sum_c CCD(c)} \quad (3)$$

Here, $CCD(c)$ is the newly introduced weight, such that $CCD(c)$ is greater when the degree of c's contribution is higher. We will elaborate the computation of this degree in section 3.

Second, concerning RSSR, as mentioned in section 1, the stronger the selectional restriction on a case of a case frame is, the less plausible an input complement satisfies that restriction. Note here that the plausibility of an interpretation of an input verb can be regarded as the plausibility that the input complements satisfy the selectional restriction associated with that interpretation. This leads us to replace $SIM(n_c, \mathcal{E}_{s,c})$ in equation (3) with $PSS(n_c, \mathcal{E}_{s,c})$, which denotes the plausibility that the case filler n_c satisfies the selectional restriction extensionary described by the example case fillers $\mathcal{E}_{s,c}$.

$$P(s) = \frac{w_s \sum_c PSS(n_c, \mathcal{E}_{s,c}) \cdot CCD(c)}{\sum_c CCD(c)} \quad (4)$$

From the assumption that $PSS(n_c, \mathcal{E}_{s,c})$ should be greater if $SIM(n_c, \mathcal{E}_{s,c})$ were greater and the relative strength of the selectional restriction described by $\mathcal{E}_{s,c}$ were less, we can derive equation (5).

$$PSS(n_c, \mathcal{E}_{s,c}) = SIM(n_c, \mathcal{E}_{s,c}) - RSSR(s, c) \quad (5)$$

Here, $RSSR(s, c)$ denotes the relative strength of the selectional restriction on a case c associated with a sense s. We will elaborate the computation of it in section 3.

3 Computation of CCD and RSSR

The degree of contribution of case to verb sense disambiguation (CCD) is computed in the following way. The degree of contribution of a case is supposed to be high if the semantic range of the example case fillers in that case is diverse in that case frame. Let us begin with the simplest case, where a certain verb has only two senses s_1 and s_2. Assume that the set of example case fillers of a case c associated with s_1 be $\mathcal{E}_{s_1,c}$ and that the set associated with s_2 be $\mathcal{E}_{s_2,c}$. Then, the degree of c's contribution to disambiguation, $CCD(c)$, is expected to be higher if the example case filler sets $\mathcal{E}_{s_1,c}$ and $\mathcal{E}_{s_2,c}$ share less elements. This can be realized by equation (6).

$$CCD(c) = \left(\frac{|\mathcal{E}_{s_1,c}| + |\mathcal{E}_{s_2,c}| - 2|\mathcal{E}_{s_1,c} \cap \mathcal{E}_{s_2,c}|}{|\mathcal{E}_{s_1,c}| + |\mathcal{E}_{s_2,c}|} \right)^{\alpha} \quad (6)$$

α is the constant for parameterizing to what extent CCD influences verb sense disambiguation. When α is larger, CCD more strongly influences the system's output. Considering the data sparseness problem, we do not distinguish two nouns X and Y in equation (6) if X and Y are similar enough as in equation (7).

$$\{X\} + \{Y\} = \{X\} \text{ if } sim(X, Y) >= 9 \quad (7)$$

For more general cases where a verb has more than two senses, say n, we average the value calculated by equation (6) for any combination of two senses as in equation (8).

$$CCD(c) = \left(\frac{1}{{}_nC_2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{|\mathcal{E}_{s_i,c}| + |\mathcal{E}_{s_j,c}| - 2|\mathcal{E}_{s_i,c} \cap \mathcal{E}_{s_j,c}|}{|\mathcal{E}_{s_i,c}| + |\mathcal{E}_{s_j,c}|} \right)^{\alpha} \quad (8)$$

Relative strength of selectional restriction (RSSR) is computed in the following way. The selectional restriction on a case of a case frame is expected to be strong if the example case fillers of the case are similar to each other. Given a set of example case fillers in a case associated with a verb sense, the strength of the selectional restriction on that case, in short SSR, can be estimated by averaging the similarity between any combination of two elements of that set. Thus, given a set $\mathcal{E}_{s,c}$ of example case fillers in a case c associated with a verb sense s, the SSR of c associated with s can be estimated by equation (9), where $\mathcal{E}^i_{s,c}$ is an i-th element of $\mathcal{E}_{s,c}$, and m is the number of elements in $\mathcal{E}_{s,c}$, i.e. $m = |\mathcal{E}_{s,c}|$.

$$SSR(s, c) = \begin{cases} \dfrac{\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} sim(\mathcal{E}^i_{s,c}, \mathcal{E}^j_{s,c})}{{}_mC_2} & \text{if } m > 1 \\ \text{maximum} & \text{otherwise} \end{cases} \quad (9)$$

In the case $m = 1$, that is, the case has only one example case filler, the SSR becomes maximum, because the selectional constraint associated with the case is highest. Following table 1, we assign 11 as maximum to SSR. The *relative* strength of selectional restriction (RSSR) of a case associated with a verb sense is estimated by the ratio of the SSR of the case to the summation of the SSRs of each case associated with the verb sense, as in equation (10).

$$RSSR(s,c) = \frac{SSR(s,c)}{\sum_i SSR(s,c_i)} \qquad (10)$$

Note that, in equation (5), while SIM is an integer, RSSR ranges in its value from 0 to 1. Therefore, RSSR is influential only when several verb senses take the same value of SIM for a given case.

4 Evaluation

This section reports the results of an experiment in which we compared the performance of the following methods:

1. Kurohashi's method: equation (1)
2. our method (considering CCD): equation (3)
3. our method (considering CCD and RSSR): equation (4)

In method 2 and 3, the influence of CCD, i.e. α in equation (8), is extremely exaggerated. In fact, the system virtually follows the procedure below:

(1) first the system relies only on the case with the highest CCD, and chooses the verb sense(s) associated with an input verb which has (have) the highest SIM value(s) for that case,

(2) if more than one verb sense is choosen in (1), the system next relies only on the case with the second highest CCD, and chooses the verb sense(s) from amongst the candidates remaining from (1), according to the SIM value(s) for that case,

(3) and so forth until the system can interpret the input verb uniquely.

Let us consider figure 1 and example sentence (2) in section 1. Suppose that the CCD of the accusative is higher than that of the nominative. In this case, first, the system relies on the SIM (see section 2) only in the accusative, for example the similarity between *syûkanshi* ("magazine") and *shinbun* ("newpaper") in the case frame of *toru* ("to subscribe"), regardless of that of other case(s), i.e. the nominative. As a result, the system chooses "to subscribe" as an interpretation of *toru*. However, if more than one verb sense associated with *toru* has the same SIM value in the accusative, then the system relies on the SIM for the nominative, and chooses the verb sense with the highest SIM for that case, in interpreting the verb.

The training/test data used in the experiment contained over one thousand simple Japansese sentences collected from a text corpus. The examples given by IPAL were also used as training data. In example-based methods such as our three methods, training data means a set of examples stored in the system. Each of the sentences in the training/test data used in our experiment consisted of one or more complement(s) followed by one of the ten verbs enumerated in table 2. For each of the ten verbs, we conducted six-fold cross validation; that is, we divided the training/test data into six equal parts, and conducted six trials in each of which a different one of the six parts was used as test data and the rest was used as training data. We shall call the former the "test set" and the latter the "training set," in each case.

When more than one interpretation of an input verb is assigned the highest plausibility score, any of the above methods will choose as its output the one that appears most frequently in training data. Therefore, the recall in each method is 100%, given that the recall is the ratio of the number of the cases where the system gives only one interpretation, to the number of inputs. Thus, in the experiment, we compared the precision of each method, which is in our case equal to the ratio of the number of correct outputs, to the number of inputs.

The performance of any corpus-based method depends on the size of training data. Given the fact that the cost of collecting training data is not at all cheap, it is important to evaluate the performance of a method in terms of the size of the training data. We first investigated how the precision of each method was improved as the training data increased. In this, we initially used only the examples given by IPAL, and progressively increased the size of the training data used, by considering an extra part of the training set (five parts of the total six data portions used) at each iteration, until finally taking all five parts in the training of our system. The number of example given by IPAL was, on average, 3.7 for each case of each case frame. The results are shown in figure 4. The x-axis denotes the ratio of the data used from the training set, to the total size of the training set.

What can be derived from figure 4 is the following. First, as more training data was considered, the precision got higher for each method. Second, consideration of CCD, i.e. contribution of case to verb sense disambiguation, improved Kurohashi's method regardless of the size of training data. Third, consideration of RSSR did not further improve the method with CCD. What should be noted is that, taking CCD into account reduced the error rate in Kurohashi's method by roughly 30%.

Table 2: Performance for each verb (*ga*: nominative, *ni*: dative, *o*: accusative, *kara*: locative, *de*: instrumental)

verb	data size	# of candidates	lower bound (%)	two highest CCD		precision (%) Kurohashi	precision (%) CCD
ataeru	136	4	66.9	*o* (0.98)	*ga* (0.88)	77.2	86.0
kakeru	160	29	25.6	*o* (0.99)	*ni* (0.98)	66.3	76.9
kuwaeru	167	5	53.9	*o* (0.99)	*ni* (0.97)	82.6	88.0
noru	126	10	45.2	*ni* (0.95)	*ga* (0.89)	82.5	81.0
osameru	108	8	25.0	*ni* (1.0)	*o* (0.98)	73.2	70.4
tsukuru	126	15	19.8	*de* (1.0)	*o* (0.98)	59.2	84.9
toru	84	29	26.2	*kara* (1.0)	*o* (0.99)	56.0	71.4
umu	90	2	81.1	*ga* (1.0)	*o* (1.0)	100	98.9
wakaru	60	5	48.3	*ga* (0.97)	*ni* (0.48)	65.0	70.0
yameru	54	2	59.3	*o* (1.0)	*ga* (0.33)	96.3	96.3
total	1111	—	43.7	—		75.2	82.4

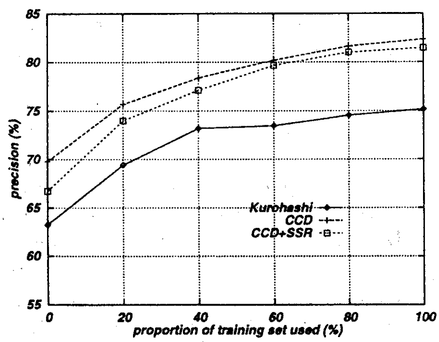

Figure 4: The precision of each method, for each size of training data

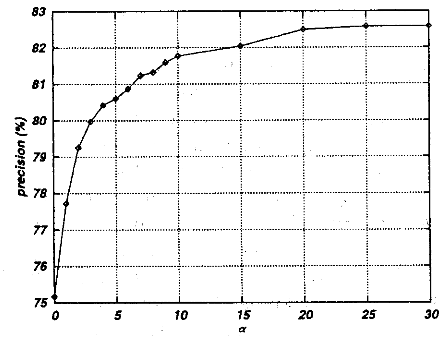

Figure 5: The relation between the degree of CCD and precision

Table 2 shows the performance for each verb gained by using the whole training set. The column of "data size" denotes the total number of sentences from the data for each verb in the training/test sets. The column of "lower bound" denotes the precision gained in a naive method such that the system always chooses the interpretation most frequently appearing in the training data [3]. The column of "two highest CCD" gives the two highest CCD values from the cases for each verb, which are calculated using only the examples described in IPAL.

Finaly, let us see to what extent we should allow CCD to influence verb sense disambiguation. Figure 5 shows the performance with the parametric constant α in equation (8) set to various values. $\alpha = 0$ corresponds with Kurohashi's method, in which CCD is never considered. As shown in figure 5, the more strong influence we allow CCD to have, the better performance we gain. This result suggets that the system can perform best when it follows the procedure mentioned at the begining of this section.

5 Discussion

Now let us consider why considering CCD is effective for verb sense disambiguation. Suppose an extreme case as shown in figure 6.

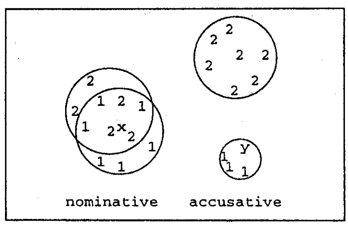

Figure 6: A case of data sparseness

In figure 6, that a symbol "1" denotes an example case filler of a certain case frame, and similary "2" denotes an example case filler of another case frame. The figure shows the distribution of example case fillers denoted by those symbols in a semantic space, where the semantic similarity between two case fillers is represented by the physical distance between two symbols. Here, suppose further that an input sentence includes two case fillers denoted by "x" and "y," for the nominative and accusative respectively. In the nominative, since "x" happens to be much closer to a "2" than any "1", "x" may be estimated to belong to the range of "2"s although "x" accually belongs to both sets of "1"s and "2"s. This case would occur if the example data were sparse. In the accusative, however, "y" would be properly estimated to belong to "1"s, even though examples did not fully cover each of the ranges of "1"s and "2"s, as these two ranges do not overlap. Note that this difference would be critical if example data were sparse. This gives us a good motivation to consider CCD in verb sense disambiguation.

6 Conclusion

In this paper, we proposed a new example-based method for verb sense disambiguation, and reported its performance through an experiment. In the experiment, our method reduced the error rate of the existing method by about 30%. This improvement was achieved by considering the degree of contribution of case to verb sense disambiguation.

For the disambiguation of English verbs, Uramoto proposed a method which integrates a rule-based method and an example-based one [11]. We expect that Uramoto's method could also be improved by introducing the notion of CCD.

As mentioned at the begining of this paper, there is another trend in corpus-based approaches, namely statistic-based approaches [1, 2, 8, 9, 12, 13]. Although many kinds of results have been reported within this class of approaches, none of researchers has presented any emprical comparison of example-based approaches and statistic-based approaches, as far as we know. It is needed to explore the condition in which each of these classes of approaches works effectively. We plan as a first step to explore how effectively each of them works given different size of training data.

The performance of our method significantly depends on the method of assigning degree of similarity to a pair of case fillers. We, at present, compute similarity in the way described in table 1. However, it is obvious that the current way needs to be revised. One possible way may be to use a statistic-based method of word clustering [10].

In our current implementation, we consider the collocation between case fillers and verbs, but ignore the combination of case fillers. Instead of a database as in figure 1, we could store a set of combinations of example case fillers, e.g. the combination of *suri* ("pickpocket") and *saifu* ("wallet"), but not that of *suri* and *otoko* ("man"). However, this way of data storage would require us to collect a much larger number of examples than the current way. This issue needs to be further investigated.

Acknoledgment

The authors would like to thank Dr. Manabu Okumura (JAIST, Japan), Dr. Michael Zock (LIMSI, France) and Mr. Timothy Baldwin (TITech, Japan) for their comments on the earlier version of this paper.

References

[1] Peter F. Brown, Stephen A. Della Pietra, and Vincent J. Della Pietra. Word-Sense Disambiguation Using Statistical Methods. In *the Proc. of ACL*, pp. 264–270, 1991.

[2] Ido Dagan and Alon Itai. Word Sense Disambiguation Using a Second Language Monolingual Corpus. *Computational Linguistics*, Vol. 20, No. 4, pp. 563–596, 1994.

[3] William Gale, Kenneth Ward Church, and David Yarowsky. Estimating Upper and Lower Bounds on the Performance of Word-Sense Disambiguation Programs. In *the Proc. of ACL*, pp. 249–256, 1992.

[4] IPA. *IPA Lexicon of the Japanese Language for computers IPAL (Basic Verbs) (in Japanese)*, 1987.

[5] Sadao Kurohashi and Makoto Nagao. A Method of Case Structure Analysis for Japanese Sentences Based on Examples in Case Frame Dictionary. *IEICE TRANSACTIONS on Information and Systems*, Vol. E77-D, No. 2, pp. 227–239, February 1994.

[6] Katashi Nagao. Recent Technical Trends in Natural Language Processing (in Japanese). *Information Processing Society of Japan*, Vol. 33, No. 7, pp. 741–745, 1992.

[7] National-Language Research Institute, editor. *Bunruigoihyo (in Japanese)*. Syuei publisher, 1964.

[8] Yoshiki Niwa and Yoshihiko Nitta. Co-occurrence vectors from corpora vs. distance vectors from dictionaries. In *the Proc. of COLING*, pp. 304–309, 1994.

[9] Hinrich Schütze. Word sense disambiguation with sublexical representations. In *Workshop Notes, Statistically-Based NLP Technieuqs, AAAI*, pp. 109–113, 1992.

[10] Takenobu Tokunaga, Makoto Iwayama, and Hozumi Tanaka. Automatic Thesaurus Construction Based on Grammatical Relations. In *the Proc. of IJCAI*, pp. 1308–1313, 1995.

[11] Naohiko Uramoto. Example-Based Word-Sense Disambiguation. *IEICE TRANSACTIONS on Information and Systems*, Vol. E77-D, No. 2, pp. 240–246, February 1994.

[12] David Yarowsky. Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora. In *the Proc. of COLING*, pp. 454–460, 1992.

[13] David Yarowsky. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. In *the Proc. of ACL*, pp. 189–196, 1995.